Finite-sample analysis of Lasso-TD

Mohammad Ghavamzadeh, Alessandro Lazaric, Rémi Munos, Matt Hoffman

Finite-Sample Analysis of Lasso-TD

Mohammad Ghavamzadeh MOHAMMAD.GHAVAMZADEH@INRIA.FR
Alessandro Lazaric ALESSANDRO.LAZARIC@INRIA.FR
Rémi Munos REMI.MUNOS@INRIA.FR
INRIA Lille - Nord Europe, Team SequeL, France

Matthew Hoffman HOFFMANM@CS.UBC.CA
Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

Abstract

In this paper, we analyze the performance of Lasso-TD, a modification of LSTD in which the projection operator is defined as a Lasso problem. We first show that Lasso-TD is guaranteed to have a unique fixed point and its algorithmic implementation coincides with the recently presented LARS-TD and LC-TD methods. We then derive two bounds on the prediction error of Lasso-TD in the Markov design setting, i.e., when the performance is evaluated on the same states used by the method. The first bound makes no assumption, but has a slow rate w.r.t. the number of samples. The second bound is under an assumption on the empirical Gram matrix, called the compatibility condition, but has an improved rate and directly relates the prediction error to the sparsity of the value function in the feature space at hand.

1. Introduction

Least-squares temporal difference (LSTD) learning (Bradtke & Barto, 1996; Boyan, 1999) is a widely used reinforcement learning (RL) algorithm for learning the value function V^π of a policy π. More precisely, LSTD tries to compute the fixed point of the operator $\Pi\mathcal{T}^\pi$, where $\mathcal{T}^\pi$ is the Bellman operator of policy π and Π is the projection operator onto a linear function space spanned by a set of d features $\{\phi_i\}_{i=1}^d$. The choice of these features has a major impact on the accuracy of the value function estimated by LSTD. The problem of finding the right space, or in other words the problem of feature selection, is an important challenge in many areas of machine learning including RL.

Appearing in *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, WA, USA, 2011.

In many situations, however, there may be no a priori way of selecting features in order to guarantee good performance. Recent approaches to value function approximation in RL instead propose to solve the problem in very high-dimensional feature spaces $\mathcal{F}$ in the hope that a good set of features lies somewhere in this basis. In practice, the fact that the number of features is often larger than the number of samples $n \leq d$ leads to the problem of *overfitting* and poor prediction performance. In order to learn from a small number of data, one should be able to select a small number of features (from the large collection of features) that are the most relevant in approximating the value function. In regression, a commonly used technique to address this issue is to solve an ℓ_1 or ℓ_2 penalized least-squares problem, called *Lasso* or *ridge*, respectively (see e.g. Hastie et al., 2001), which provides a way to regularize the objective function in order to overcome the overfitting problem. Among these two methods, Lasso is of particular interest, because the geometric properties of the ℓ_1-penalty encourage *sparse* solutions, i.e., solutions that can be expressed in terms of the linear combination of a small number of features. This is exactly the type of solution we would hope for when moving to high-dimensional spaces.

In value function approximation in RL, however, the objective is not to recover a target function given its noisy observations, but is instead to approximate the fixed-point of the Bellman operator given sample trajectories. This creates some difficulties in applying Lasso and ridge to this problem. Despite these difficulties, both ℓ_1 and ℓ_2 regularizations have been previously studied in value function approximation in RL. Farahmand et al. presented several such algorithms wherein ℓ_2-regularization was added to LSTD and modified Bellman residual minimization (Farahmand et al., 2008), to fitted Q-iteration (Farahmand et al., 2009), and finite-sample performance bounds for these algorithms were proved. There has also been al-

gorithmic work on adding ℓ_1-penalties to the TD (Loth et al., 2007), LSTD (Kolter & Ng, 2009; Johns et al., 2010), and linear programming (Petrik et al., 2010) algorithms.

In the work by Kolter & Ng (2009) and Johns et al. (2010), the idea is to find a fixed point solution to an ℓ_1-penalized LSTD problem. The experimental results reported in these papers are interesting and ask for theoretical analysis. In this paper, we consider a modification of LSTD, which incorporates an ℓ_1-penalty in its projection operator. We call this problem *Lasso-TD* and analyze it in the setting of *pathwise LSTD* (Lazaric et al., 2010). Our analysis may apply to any method which solves for the underlying fixed-point, including the recently proposed LARS-TD (Kolter & Ng, 2009) and LC-TD (Johns et al., 2010) algorithms. In this work, we show that a Lasso-TD solution $\tilde{\alpha}$ always exists and its corresponding value function approximation $\tilde{v} = f_{\tilde{\alpha}}$ is unique. Beyond this, the main objective of this paper is to investigate which "nice" properties of the ℓ_1-penalized regression carry forward into the fixed-point setting. As a result, we also provide two bounds on the approximation error $\|v - f_{\tilde{\alpha}}\|_n$, where v is the true value function and $\|\cdot\|_n$ denotes the empirical norm at the states used by the algorithm. The first of these bounds makes no assumptions and has estimation error which scales with the ℓ_1-norm of the best approximation at a rate of $n^{-1/4}$. Under the compatibility condition (van de Geer & Bühlmann, 2009) on the features (more precisely on the empirical Gram matrix $\frac{1}{n}\Phi^\top\Phi$), we prove a more refined bound which scales with the ℓ_0-norm of the best approximation (i.e., the actual sparsity) and with a better rate $n^{-1/2}$. Our results indicate that if the value function can be well approximated with a small number of relevant features, then the number of samples n required to learn a good approximation of it only scales with the number of relevant features.

2. Preliminaries

For a measurable space with domain $\mathcal{X}$, we let $\mathcal{S}(\mathcal{X})$ and $\mathcal{B}(\mathcal{X}; L)$ denote the set of probability measures over $\mathcal{X}$, and the space of bounded measurable functions with domain $\mathcal{X}$ and bound $0 < L < \infty$, respectively. For a measure $\rho \in \mathcal{S}(\mathcal{X})$ and a measurable function $f : \mathcal{X} \to \mathbb{R}$, we define the $\ell_2(\rho)$-norm of f, $||f||_\rho$, and for a set of n states $X_1, \ldots, X_n \in \mathcal{X}$, we define the empirical norm $||f||_n$ as

$$||f||_\rho^2 = \int f(x)^2 \rho(dx) \qquad \text{and} \qquad ||f||_n^2 = \frac{1}{n}\sum_{t=1}^n f(X_t)^2.$$

We consider the standard RL framework (Sutton & Barto, 1998) in which a learning agent interacts with a stochastic environment by following a policy π and this interaction is modeled as a discrete-time discounted Markov chain (MC). A discounted MC is a tuple $\mathcal{M}^\pi = \langle \mathcal{X}, R^\pi, P^\pi, \gamma \rangle$, where the state space $\mathcal{X}$ is a subset of a Euclidean space, the reward function $R^\pi : \mathcal{X} \to \mathbb{R}$ is uniformly bounded by $R_{\max}$, the transition kernel P^π is such that for all $x \in \mathcal{X}$, $P^\pi(\cdot|x)$ is a distribution over $\mathcal{X}$, and $\gamma \in (0,1)$ is a discount factor. The value function of a policy π, V^π, is the unique fixed-point of the Bellman operator $\mathcal{T}^\pi : \mathcal{B}(\mathcal{X}; V_{\max} = \frac{R_{\max}}{1-\gamma}) \to \mathcal{B}(\mathcal{X}; V_{\max})$ defined by[1]

$$(\mathcal{T}^\pi V)(x) = R^\pi(x) + \gamma \int_{\mathcal{X}} P^\pi(dy|x) V(y).$$

To approximate the value function V^π, we use a linear approximation architecture with parameter $\alpha \in \mathbb{R}^d$ and basis functions $\varphi_j \in \mathcal{B}(\mathcal{X}; L)$, $j = 1, \ldots, d$. We denote by $\phi : \mathcal{X} \to \mathbb{R}^d$, $\phi(\cdot) = \big(\varphi_1(\cdot), \ldots, \varphi_d(\cdot)\big)^\top$ the feature vector, and by $\mathcal{F}$ the linear function space spanned by the basis functions φ_j.

Let $(X_1, \ldots, X_n)$ be a sample path (or trajectory) of size n generated by the Markov chain $\mathcal{M}$. Let $v \in \mathbb{R}^n$ and $r \in \mathbb{R}^n$ such that $v_t = V^\pi(X_t)$ and $r_t = R(X_t)$ be the value vector and the reward vector, respectively. Also, let $\Phi = [\phi(X_1)^\top; \ldots; \phi(X_n)^\top]$ be the feature matrix defined at the states, and $\mathcal{F}_n = \{f_\alpha = \Phi\alpha : \alpha \in \mathbb{R}^d\} \subset \mathbb{R}^n$ be the corresponding vector space.

3. Lasso-TD

Lasso-TD takes a single trajectory $\{X_t\}_{t=1}^n$ of size n generated by the Markov chain and the observed rewards $\{R(X_t)\}_{t=1}^n$ as input, and computes the fixed point $\tilde{v} = \widehat{\Pi}_\lambda \widehat{\mathcal{T}} \tilde{v}$ similar to LSTD. As in LSTD, $\widehat{\mathcal{T}} : \mathbb{R}^n \to \mathbb{R}^n$ is the *pathwise emperical Bellman operator* (Lazaric et al., 2010) defined as

$$(\widehat{\mathcal{T}} y)_t = \begin{cases} r_t + \gamma y_{t+1} & 1 \le t < n, \\ r_t & t = n. \end{cases}$$

Lasso-TD departs from LSTD in the definition of the projection operator. Here $\widehat{\Pi}_\lambda : \mathbb{R}^n \to \mathcal{F}_n$ is the ℓ_1-*penalized projection* defined as

$$\widehat{\Pi}_\lambda y = \Phi\hat{\alpha} \quad \text{such that} \quad \hat{\alpha} = \arg\min_{\alpha \in \mathbb{R}^d} ||y - f_\alpha||_n^2 + \lambda ||\alpha||_1 .$$

In order to show the existence and uniqueness of the fixed point $\tilde{v}$, we show that $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$ is a γ-contraction w.r.t. the ℓ_2-norm. Lazaric et al. (2010) showed that $\widehat{\mathcal{T}}$ is a γ-contraction mapping in ℓ_2-norm, and Lemma 1 below proves that $\widehat{\Pi}_\lambda$ is a non-expansive mapping in

[1] To simplify the notation, we remove the dependency to the policy π and use $\mathcal{M}$, R, P, V, and $\mathcal{T}$ instead of $\mathcal{M}^\pi$, R^π, P^π, V^π, and $\mathcal{T}^\pi$ throughout the paper.

the same norm. As a consequence, $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$ is a contraction mapping and from the Banach fixed point theorem, there exists a unique fixed point $\tilde{v} \in \mathcal{F}_n$ of $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$. We define the Lasso-TD solution $\tilde{\alpha}$ as

$$\tilde{\alpha} = \arg\min_{\alpha \in \mathbb{R}^d} ||\widehat{\mathcal{T}}\tilde{v} - f_\alpha||_n^2 + \lambda ||\alpha||_1 \ . \tag{1}$$

From the definition of the projection $\widehat{\Pi}_\lambda$ we have that $f_{\tilde{\alpha}} = \Phi\tilde{\alpha} = \widehat{\Pi}_\lambda \widehat{\mathcal{T}}\tilde{v}$, and from the fact that $\tilde{v}$ is the fixed point of $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$, we deduce that $\tilde{\alpha}$ is solution to the ℓ_1-penalized fixed-point equation:

$$\tilde{\alpha} = \arg\min_{\alpha \in \mathbb{R}^d} \|\widehat{\mathcal{T}} f_{\tilde{\alpha}} - f_\alpha\|_n^2 + \lambda \|\alpha\|_1 \ . \tag{2}$$

This latter formulation is the one used in Kolter & Ng (2009) and Johns et al. (2010). By introducing a contraction mapping in the function space $\mathcal{F}_n$ we are able to prove the existence of the fixed-point $\tilde{v} \in \mathcal{F}_n$, and deduce that the parameter $\tilde{\alpha}$ (as defined by Eq. 1) is the solution to Eq. 2, which is an ℓ_1-penalized fixed-point formulation in the parameter space. Note that although $\tilde{v}$ is unique, the solution $\tilde{\alpha}$ of Eq. 2 may not be unique (if there are several minimizers to the problem in Eq. 1), but any solution α to Eq. 2 is such that $\Phi\alpha$ is a fixed point of $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$, thus $\Phi\alpha = \tilde{v}$. From an algorithmic perspective, one may use the formulation in Eq. 2 in order to compute $\tilde{\alpha}$ (see e.g. Kolter & Ng, 2009; Johns et al., 2010). We will now prove that the projection operator is a non-expansion.

Lemma 1. *For any $x, y \in \mathbb{R}^n$, $\widehat{\Pi}_\lambda$ satisfies*

$$\|\widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y\|_n^2 \le \|x - y\|_n^2 - \|x - y - (\widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y)\|_n^2. \tag{3}$$

Proof. We prove the result for a general proximal operator. Let $g : \mathbb{R}^d \to \mathbb{R}$ be any convex function (e.g., $g(\alpha) = \lambda\|\alpha\|_1$). For any $y \in \mathbb{R}^n$, we define

$$\alpha_y = \arg\min_{\alpha \in \mathbb{R}^d} \|\Phi\alpha - y\|_n^2 + g(\alpha),$$

and write $\widehat{\Pi}_\lambda y = \Phi\alpha_y$. Since α_y is a minimum,

$$0 \in \partial\big(\|y - \Phi\cdot\|_n^2 + g(\cdot)\big)(\alpha_y) = -\frac{2}{n}\Phi^\top(y - \Phi\alpha_y) + \partial g(\alpha_y),$$

and thus $\frac{2}{n}\Phi^\top(y - \Phi\alpha_y) \in \partial g(\alpha_y)$. From the definition of sub-gradient, for any $z \in \mathbb{R}^d$, we have

$$\big\langle z - \alpha_y, \frac{2}{n}\Phi^\top(y - \Phi\alpha_y)\big\rangle \le g(z) - g(\alpha_y).$$

By choosing $z = \alpha_x$, we obtain $\big\langle \alpha_x - \alpha_y, \frac{2}{n}\Phi^\top(y - \Phi\alpha_y)\big\rangle \le g(\alpha_x) - g(\alpha_y)$, which may be rewritten as

$$\frac{2}{n}\big\langle \widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y, y - \widehat{\Pi}_\lambda y\big\rangle \le g(\alpha_x) - g(\alpha_y).$$

With a similar reasoning, we may deduce

$$\frac{2}{n}\big\langle \widehat{\Pi}_\lambda y - \widehat{\Pi}_\lambda x, x - \widehat{\Pi}_\lambda x\big\rangle \le g(\alpha_y) - g(\alpha_x).$$

By adding these inequalities, we obtain

$$\|\widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y\|_n^2 \le \langle \widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y, x - y\rangle.$$

The claim follows using the fact that

$$\|x - y - (\widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y)\|_n^2 = \|x - y\|_n^2 + \|\widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y\|_n^2 - 2\langle \widehat{\Pi}_\lambda x - \widehat{\Pi}_\lambda y, x - y\rangle. \qquad \square$$

From Lemma 1, we deduce that $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$ is a contraction mapping w.r.t. the ℓ_2-norm, which guarantees the existence and uniqueness of the fixed point $\tilde{v}$ of $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$. Note that we did not require any assumptions in order to show the existence and uniqueness of this solution, however, additional assumptions may be necessary in order to find these solutions algorithmically (Kolter & Ng, 2009; Johns et al., 2010).

4. Markov Design Bounds

In Section 3, we defined Lasso-TD and showed that the operator $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$ always has a unique fixed point $\Phi\tilde{\alpha}$. In this section, we derive two bounds for the performance of $\Phi\tilde{\alpha}$, i.e., the performance of $f_{\tilde{\alpha}}$ evaluated at the states of the trajectory used by the algorithm. In Section 4.1, we derive a performance bound that has an estimation error with the rate $O(||\alpha||_1 n^{-1/4})$. Although this bound has a poor rate, it does not require any assumption on the empirical Gram matrix $\frac{1}{n}\Phi^\top\Phi$. In Section 4.2, we derive a bound with a better rate $O(||\alpha||_0 n^{-1/2})$, where this improvement is at the cost of introducing an additional assumption on the empirical Gram matrix, called the *compatibility condition*.

4.1. ℓ_1-Oracle Inequality

The main theorem of this section shows a natural connection between the prediction performance of Lasso-TD (which is an ℓ_1-regularized fixed point problem) and the ℓ_1-norm of the best approximation of v in $\mathcal{F}_n$.

Theorem 1. *Let $\delta > 0$. Let $\{X_t\}_{t=1}^n$ be a trajectory generated by the Markov chain, and let $v, \Phi\tilde{\alpha} \in \mathbb{R}^n$ be vectors whose components are the value function and the Lasso-TD prediction at $\{X_t\}_{t=1}^n$, respectively. Then, with probability $1 - \delta$ (w.r.t. the trajectory), we have*

$$||v - f_{\tilde{\alpha}}||_n \le \frac{1}{1-\gamma} \inf_{\alpha \in \mathbb{R}^d} \Bigg[||v - f_\alpha||_n + \sqrt{12\gamma V_{\max} L ||\alpha||_1} \left(\left(\frac{2\log(2d/\delta)}{n} \right)^{1/4} + \frac{1}{\sqrt{2n}} \right) \Bigg]. \tag{4}$$

Remark 1. We should first emphasize that this bound makes no assumption on either the state distribution or the empirical Gram matrix (related to

the correlation of the features). The bound consists of an approximation error term (i.e., the ℓ_2 distance between v and the function f_α) and an estimation error term that scales with the ℓ_1-norm $\|\alpha\|_1$ at the rate $n^{-1/4}$. Intuitively, this means that if there exists a good approximation $f_{\bar{\alpha}} \in \mathcal{F}_n$ of the value function that has small ℓ_1-norm, i.e., $||\bar{\alpha}||_1$ is small, then the approximation $f_{\hat{\alpha}}$ obtained by Lasso-TD will perform almost as well as $f_{\bar{\alpha}}$. For this reason, we refer to this bound as an ℓ_1-oracle inequality. This result perfectly matches the results for ℓ_1-regularized methods for regression in the deterministic design setting where the prediction error of the Lasso solution is related to the ℓ_1-norm of the best approximation of the target function in $\mathcal{F}_n$ (see e.g., Massart & Meynet 2010). Finally, although the properties of the ℓ_1-penalty tend to produce "sparse" solutions and often behave similar to the ℓ_0-penalty, we may not deduce such behavior from this bound. However in Section 4.2, we will show that under suitable conditions the Lasso-TD solution has a strong connection with the sparsity of the functions that approximate v in $\mathcal{F}_n$.

Remark 2. (Lasso-TD vs. LSTD) We may also compare the Lasso-TD bound to the Markov design LSTD bound reported in Lazaric et al. (2010). To ease the comparison, we report simplified versions of both bounds where $\tilde{O}$ hides constants, logarithmic terms in δ, and dominated terms. Let $\alpha^* = \arg\min_{\alpha\in\mathbb{R}^d} ||v - f_\alpha||_n$ and let α_L be the LSTD solution. The Lasso-TD bound in Eq. 4 may be written as

$$||v - f_{\hat{\alpha}}||_n \le \frac{1}{1-\gamma}\left[||v - f_{\alpha^*}||_n + \tilde{O}\left(\left(\frac{||\alpha^*||_1^2 \log d}{n}\right)^{\frac{1}{4}}\right)\right],$$

while the performance of LSTD can be bounded as

$$||v - f_{\alpha_L}||_n \le \frac{1}{\sqrt{1-\gamma^2}}||v - f_{\alpha^*}||_n + \frac{\gamma}{1-\gamma}\tilde{O}\left(\sqrt{\frac{d}{n\nu_n}}\right),$$

where ν_n is the smallest strictly positive eigenvalue of the empirical Gram matrix. Although in Lasso-TD the constant $(1-\gamma)^{-1}$ in front of the approximation error is larger than $(1-\gamma^2)^{-1/2}$, both bounds share the same dependency on the accuracy of f_{α^*}, the best approximation of the value function in the vector space $\mathcal{F}_n$. On the other hand, the estimation errors display significantly different behaviors. The estimation error of LSTD has a better rate $n^{-1/2}$ compared to the rate $n^{-1/4}$ for Lasso-TD. However, Lasso-TD attains a much smaller estimation error whenever the dimensionality d is of the same order as (or larger than) the number of samples. Similar to regression, in this case LSTD tends to overfit the noise in the data, while Lasso-TD performs better as a regularization method. In fact, in LSTD the estimation error increases as $d^{1/2}$, while Lasso-TD has a milder dependency of order $||\alpha^*||_1 \log d$. Finally, we note that the bound for LSTD reported in Lazaric et al. (2010) has an additional dependency on $\nu_n^{-1/2}$ which does not appear in the bound for Lasso-TD.

Remark 3. (Lasso-TD vs. LSTD-RP) Another interesting comparison is to the LSTD with random projections (LSTD-RP) algorithm introduced in Ghavamzadeh et al. (2010). LSTD-RP first generates a new function space $\mathcal{G}$ spanned by d' (with $d' \ll d$) features obtained as linear random combinations of the features φ_i of $\mathcal{F}$ and then runs standard LSTD on it. Let α_{RP} be the solution returned by LSTD-RP. After optimizing the dimensionality d' of the randomly generated linear space as suggested in Ghavamzadeh et al. (2010), LSTD-RP achieves the following performance bound

$$||v - f_{\alpha_{RP}}||_n \le \frac{1}{\sqrt{1-\gamma^2}}||v - f_{\alpha^*}||_n + \frac{\gamma}{1-\gamma}\tilde{O}\left(\left(\frac{||\alpha^*||_2^2 \log n}{n\nu_n}\right)^{1/4}\right).$$

Unlike LSTD, the estimation error in both Lasso-TD and LSTD-RP has the same decreasing rate w.r.t. the number of samples n. The main difference here is that Lasso-TD scales with $||\alpha^*||_1$, while $||\alpha^*||_2$ appears in the LSTD-RP bound. This difference is consistent with the definition of the two algorithms as ℓ_1 and ℓ_2 regularization methods. Since $||\alpha^*||_2 \le ||\alpha^*||_1$, LSTD-RP bound is in general tighter than Lasso-TD's. However, in Thm. 2 we show that under certain assumptions, the Lasso-TD bound may be refined and its estimation error would depend on the ℓ_0-norm of α^* (the sparsity of α^*) and would have a better rate $n^{-1/2}$.

In order to prove Thm. 1, we first state and prove a Lemma about the *Markov design* model (Lazaric et al., 2010). The model of regression with Markov design is a regression problem where the data $\{(X_t, Y_t)\}_{t=1}^n$ are generated such that $\{X_t\}_{t=1}^n$ is a trajectory sampled from a Markov chain, $Y_t = f(X_t) + \xi_t$ consists of noisy function evaluations, and f is the target function. Each noise term ξ_t is bounded and adapted to the filtration generated by $X_{1:t+1}$, and have zero-mean conditioned on $X_{1:t}$, i.e.,

$$|\xi_t| \le C \qquad \text{and} \qquad \mathbb{E}[\xi_t | X_1, \dots, X_t] = 0\,, \tag{5}$$

The following lemma provides a bound on the performance of Lasso for the Markov design setting that has a slow rate, but does not require any assumptions.

Lemma 2. *We consider the model of regression with Markov design. Let $f_{\hat{\alpha}} \in \mathcal{F}_n$ be the Lasso prediction of the (noisy) values $Y = \{Y_t\}_1^n$, i.e., $f_{\hat{\alpha}} = \Phi\hat{\alpha} = \widehat{\Pi}_\lambda Y$. For any $\delta > 0$, with probability at least $1-\delta$, we have*

$$||f - f_{\hat{\alpha}}||_n^2 \le ||f - f_\alpha||_n^2 + 6CL||\alpha||_1 \sqrt{\frac{2\log(2d/\delta)}{n}} \,,$$

for any $\Phi\alpha \in \mathcal{F}_n$.

Proof of Lemma 2. We define $\xi \in \mathbb{R}^n$ to be the vector with components ξ_t. From the definition of $\hat{\alpha}$ we have

$$||Y - f_{\hat{\alpha}}||_n^2 + \lambda||\hat{\alpha}||_1 \le ||Y - f_\alpha||_n^2 + \lambda||\alpha||_1 \,. \tag{6}$$

We may write $||Y - f_{\hat{\alpha}}||_n^2 - ||Y - f_\alpha||_n^2$ as

$$||Y - f_{\hat{\alpha}}||_n^2 - ||Y - f_\alpha||_n^2 = ||Y - f||_n^2 + 2\langle Y - f, f - f_{\hat{\alpha}}\rangle_n +$$
$$||f - f_{\hat{\alpha}}||_n^2 - \Big(||Y - f||_n^2 + 2\langle Y - f, f - f_\alpha\rangle_n + ||f - f_\alpha||_n^2\Big)$$
$$= ||f - f_{\hat{\alpha}}||_n^2 - ||f - f_\alpha||_n^2 + 2\langle \xi, f_\alpha - f_{\hat{\alpha}}\rangle_n \,,$$

and thus, Eq. 6 may be written as

$$||f - f_{\hat{\alpha}}||_n^2 + \lambda||\hat{\alpha}||_1 \le ||f - f_\alpha||_n^2 + \lambda||\alpha||_1 + 2\langle \xi, f_{\hat{\alpha}} - f_\alpha\rangle_n \,. \tag{7}$$

The term $2\langle \xi, f_{\hat{\alpha}} - f_\alpha\rangle_n$ in Eq. 7 may be bounded as

$$2\langle \xi, f_{\hat{\alpha}} - f_\alpha\rangle_n \le \frac{2}{n} \max_{i=1,\dots,d} \Big|\sum_{t=1}^n \xi_t \varphi_i(X_t)\Big| \, ||\hat{\alpha} - \alpha||_1 \,.$$

We now define the event

$$\mathcal{E}_1 = \Big\{ \frac{2}{n} \max_{i=1,\dots,d} \Big|\sum_{t=1}^n \xi_t \varphi_i(X_t)\Big| \le \lambda_0 \Big\} \,. \tag{8}$$

Under $\mathcal{E}_1$ we have

$$2\langle \xi, f_{\hat{\alpha}} - f_\alpha\rangle_n \le \lambda_0||\hat{\alpha} - \alpha||_1 \le \lambda_0||\hat{\alpha}||_1 + \lambda_0||\alpha||_1 \,,$$

and thus using Eq. 7, we obtain

$$||f - f_{\hat{\alpha}}||_n^2 + (\lambda - \lambda_0)||\hat{\alpha}||_1 \le ||f - f_\alpha||_n^2 + (\lambda + \lambda_0)||\alpha||_1 \,. \tag{9}$$

For any $\lambda \ge 2\lambda_0$, Eq. 9 may be written as

$$||f - f_{\hat{\alpha}}||_n^2 + \frac{\lambda}{2}||\hat{\alpha}||_1 \le ||f - f_\alpha||_n^2 + \frac{3\lambda}{2}||\alpha||_1 \,. \tag{10}$$

From the conditions on the noise in the Markov design setting, we have that for any $i = 1, \dots, d$

$$\mathbb{E}[\xi_t \varphi_i(X_t)|X_1, \dots, X_t] = \varphi_i(X_t)\mathbb{E}[\xi_t|X_1, \dots, X_t] = 0 \,,$$

and since $\xi_t \varphi_i(X_t)$ is adapted to the filtration generated by $X_1, \dots, X_{t+1}$, it is a martingale difference sequence w.r.t. that filtration. Hence by applying Azuma's inequality, we may deduce that with probability $1 - \delta$

$$\Big|\sum_{t=1}^n \xi_t \varphi_i(X_t)\Big| \le CL\sqrt{2n\log(2/\delta)} \,,$$

where we used the fact that $|\xi_t \varphi_i(X_t)| \le CL$ for any i and t. By union bound over all features, we have that with probability $1 - \delta$

$$\max_{i=1,\dots,d} \Big|\sum_{t=1}^n \xi_t \varphi_i(X_t)\Big| \le CL\sqrt{2n\log(2d/\delta)} \,, \tag{11}$$

and thus with probability $1 - \delta$

$$\frac{2}{n} \max_{i=1,\dots,d} \Big|\sum_{t=1}^n \xi_t \varphi_i(X_t)\Big| \le 2CL\sqrt{\frac{2\log(2d/\delta)}{n}} \,.$$

The result follows by setting $\lambda_0 = 2CL\sqrt{\frac{2\log(2d/\delta)}{n}}$, $\lambda = 2\lambda_0$, and plugging this value of λ in Eq. 10. □

Proof of Theorem 1. **Step 1:** Using the triangle inequality, we have

$$||v - f_{\tilde{\alpha}}||_n \le ||v - \widehat{\Pi}_\lambda \widehat{\mathcal{T}} v||_n + ||\widehat{\Pi}_\lambda \widehat{\mathcal{T}} v - f_{\tilde{\alpha}}||_n \,. \tag{12}$$

From the γ-contraction of $\widehat{\Pi}_\lambda \widehat{\mathcal{T}}$, and the fact that $f_{\tilde{\alpha}}$ is its unique fixed point, we obtain

$$||\widehat{\Pi}_\lambda \widehat{\mathcal{T}} v - f_{\tilde{\alpha}}||_n = ||\widehat{\Pi}_\lambda \widehat{\mathcal{T}} v - \widehat{\Pi}_\lambda \widehat{\mathcal{T}} f_{\tilde{\alpha}}||_n \le \gamma||v - f_{\tilde{\alpha}}||_n \,. \tag{13}$$

Thus from Eq. 12 and 13, we have

$$||v - f_{\tilde{\alpha}}||_n \le \frac{1}{1 - \gamma}||v - \widehat{\Pi}_\lambda \widehat{\mathcal{T}} v||_n \,. \tag{14}$$

Step 2: We now provide a high probability bound for $||v - \widehat{\Pi}_\lambda \widehat{\mathcal{T}} v||_n$. This is a consequence of Lemma 2 applied to the vectors $Y = \widehat{\mathcal{T}} v$ and $\big[f(X_t)\big]_{t=1}^n = v$. Since v is the value function at the points $\{X_t\}_{t=1}^n$, from the definition of the pathwise Bellman operator, we have that for $1 \le t \le n-1$, $\xi_t = \gamma\big[V(X_t) - \int P(dy|X_t)V(y)\big]$, and $\xi_n = -\gamma \int P(dy|X_n)V(y)$. Thus, Eq. 5 holds for $1 \le t \le n-1$. Here we may choose $C = 2\gamma V_{\max}$ for a bound on $\{\xi_t\}_{t=1}^{n-1}$, and $C = \gamma V_{\max}$ for a bound on ξ_n. Azuma's inequality may be applied only to the sequence of $n-1$ terms, thus instead of Eq. 11, we obtain

$$\max_{i=1,\dots,d} \Big|\sum_{t=1}^n \xi_t \varphi_i(X_t)\Big| \le \gamma V_{\max} L\big(2\sqrt{2n\log(2d/\delta)} + 1\big) \,,$$

with probability $1 - \delta$. Setting $\lambda_0 = 2\gamma V_{\max} L\big(2\sqrt{\frac{2\log(2d/\delta)}{n}} + \frac{1}{n}\big)$, $\lambda = 2\lambda_0$, and plugging in this value of λ in Eq. 10, we deduce that with probability $1 - \delta$, we have

$$||v - \widehat{\Pi}_\lambda \widehat{\mathcal{T}} v||_n^2 \le ||v - f_\alpha||_n^2 \tag{15}$$
$$+ 12\gamma V_{\max} L||\alpha||_1\Big(\sqrt{\frac{2\log(2d/\delta)}{n}} + \frac{1}{2n}\Big).$$

The claim follows by taking the square-root of Eq. 15 and combining it with Eq. 14. □

4.2. Sparsity Oracle Inequality

In Section 4.1, we derived a bound on the prediction performance of the Lasso-TD solution which does not require any assumption on the existence of the stationary distribution of the policy under evaluation and on the empirical Gram matrix. However, as discussed in Remark 2, the estimation error in Thm. 1 has a poor rate (compared to LSTD, $n^{-1/4}$ instead of $n^{-1/2}$) and scales with the ℓ_1-norm of α, which is not directly related to the sparsity of α. In this section, we derive an improved bound whose estimation error depends on $||\alpha||_0$ and decreases as $n^{-1/2}$ at the cost of an extra assumption on the empirical Gram matrix $\frac{1}{n}\Phi^\top\Phi$.

Prediction bounds in the deterministic design setting for Lasso in linear models may be derived under a variety of different assumptions on the empirical Gram matrix. The most common assumption is the celebrated restricted isometry property (RIP) (see e.g., Candes & Tao 2005) which guarantees that the feature matrix restricted on an index set S approximately preserves the norm of any vector. Although sufficient to prove both prediction and reconstruction properties of Lasso, the RIP assumption is rarely satisfied in practice. van de Geer & Bühlmann (2009) report a thorough analysis of a number of different sufficient condition for sparsity oracle inequalities (i.e., performance bounds depending on the sparsity of the problem). The *compatibility condition* is one of the weakest of these assumptions, and thus, may be satisfied by a fairly general class of Gram matrices (van de Geer & Bühlmann, 2009).

Before stating the main result of this section, we introduce some notations and define a condition on the empirical Gram matrix $\frac{1}{n}\Phi^\top\Phi$. Let $S \subseteq \{1,\dots,d\}$ be an index set, $s = |S|$ be its cardinality, and $S^c = \{1,\dots,d\} - S$ be its complement. For a vector $\alpha \in \mathbb{R}^d$ and an index set $S \subseteq \{1,\dots,d\}$, we denote by $\alpha_S \in \mathbb{R}^d$ the vector with non-zero entries in the index set S, i.e., $\alpha_S^{(i)} = \alpha_i \mathbb{I}\{i \in S\}$. For a vector $\alpha \in \mathbb{R}^d$, we denote by $S_\alpha = \{i : \alpha_i \neq 0\}$ the set of indices of the non-zero entries, or the *active set*, of α, $s_\alpha = |S_\alpha|$ the cardinality of S_α or the *sparsity index* of α, and $S_\alpha^c = \{i : \alpha_i = 0\}$ the complement of S_α. Note that from the definition of S_α, we have $\alpha_{S_\alpha^c} = \vec{0}$.

Definition 1 (Compatibility Condition). *Let $S \subseteq \{1,\dots,d\}$ be an index set and $\mathcal{R}(K,S) = \{\alpha \in \mathbb{R}^d : ||\alpha_{S^c}||_1 \leq K||\alpha_S||_1 \neq 0\}$ be a restriction set. We call*

$$\psi^2(K,S) = \min\left\{\frac{s||f_\alpha||_n^2}{||\alpha_S||_1^2} : \alpha \in \mathcal{R}(K,S)\right\} \qquad (16)$$

the (K,S)-restricted ℓ_1-eigenvalue of the empirical Gram matrix. The empirical Gram matrix satisfies the (K,S)-compatibility condition if $\psi(K,S) > 0$.

We are now ready to derive a better bound for the prediction performance of Lasso-TD.

Theorem 2. *Let $\{X_t\}_{t=1}^n$ be a trajectory generated by the Markov chain, and $v, \Phi\tilde{\alpha} \in \mathbb{R}^n$ be the vectors whose components are the value function and the Lasso-TD prediction at $\{X_t\}_{t=1}^n$, respectively. For any $\delta > 0$, with probability $1-\delta$, we have*

$$||v - f_{\tilde{\alpha}}||_n \leq \frac{1}{1-\gamma}\inf_{\alpha\in U}\left[||v - f_\alpha||_n + \frac{12\gamma V_{\max} L\sqrt{s}}{\psi}\left(\sqrt{\frac{2\log(2d/\delta)}{n}} + \frac{1}{2n}\right)\right], \qquad (17)$$

for any $\alpha \in \mathbb{R}^d$ such that the empirical Gram matrix $\frac{1}{n}\Phi^\top\Phi$ satisfies the $(3, S_\alpha)$-compatibility condition, where $U \subseteq \mathbb{R}^d$ is the set of such α's.

Remark 1. (Sparsity Oracle Inequality) The bound of Eq. 17 shares the same structure as the bound in Eq. 4, where the prediction error is divided into an approximation error and an estimation error terms. The main difference is that the estimation error improved from $\tilde{O}\big((||\alpha||_1^2 \log d/n)^{1/4}\big)$ to $\tilde{O}\big((s\log d/\psi n)^{1/2}\big)$, where $s = ||\alpha||_0$. It is interesting to note that up to a logarithmic factor $\log d$ and ignoring the dependency on ψ and ν_n, this is the same performance LSTD would obtain using the features in the set S_α instead of the whole function space $\mathcal{F}$. This type of bounds is usually referred to as *sparsity oracle inequalities*. In fact, Eq. 17 shows that if there exists an s-sparse α with small approximation error (e.g., the minimizer of the RHS of the bound), Lasso-TD is able to take advantage of its sparsity and obtain a prediction performance which is approximately equivalent to the one achieved by searching directly in the space with dimensionality s. Furthermore, this performance bound also shows the effectiveness of Lasso-TD in solving high-dimensional problems where the number of features d is larger than the number of samples n. In fact, with a fixed number of samples n, as d increases, the performance of LSTD becomes worse and worse, while Lasso-TD attains an almost constant prediction accuracy ($\tilde{O}(\log d)$), because the number of relevant features s does not change with d. Finally, this result supports the claim that the fixed point method Lasso-TD has the same properties as Lasso in regression.

Remark 2. (Sparse Value Functions) As discussed in Remark 1, in order for Lasso-TD to be effective, a sparse approximation of the value function must exist. While in regression this assumption is natural

in many domains, in RL we need to characterize the control problems which are likely to generate sparse value functions in a given feature space. Although a full analysis of this problem is beyond the scope of this paper, in the following we try to present some characteristics which are sufficient to expect Lasso-TD to perform well. Let us consider the class of continuous MDPs with finite actions, smooth dynamics (i.e., smooth stochastic transition kernel), and smooth reward function. In this case, for any policy π, the corresponding value function V^π is piecewise smooth.[2] From approximation theory, we know that piecewise-smooth functions can be accurately approximated by localized features, such as wavelets (see e.g., Mallat 1999). As a result, if we define $\mathcal{F}$ as the linear function space spanned by a set of d wavelets, any value function is likely to be well-approximated by functions in $\mathcal{F}$ with a relatively small number of non-zero coefficients. In this case, even if d is large so as to guarantee small approximation errors, we expect Lasso-TD to require few samples (i.e., n be of the same order as the sparsity of the best solution) to learn an accurate approximation of the value function at hand.

Remark 3. (Lasso-TD vs. LSTD-RP) In Remark 3 of Section 4.1, we mentioned that LSTD-RP may achieve a better performance than Lasso-TD in the general no-assumption case. On the other hand, when the compatibility assumption holds, Lasso-TD achieves an estimation error of order $\tilde{O}\big(\sqrt{(s \log d)/(\psi n)}\big)$ which has a better decreasing rate than LSTD-RP and displays a direct connection with the sparsity s of α. On the other hand, LSTD-RP does not take advantage of the possible sparsity of α and provides bounds in terms of the ℓ_2-norm of α, which may be larger than s.

Remark 4. As mentioned at the beginning of this section, many different assumptions have been proposed to derive sparsity oracle inequalities for prediction in regression, among which compatibility is the weakest sufficient condition. In Thm. 2, we showed that similar to regression, the compatibility condition is a sufficient assumption to prove a sparsity oracle inequality for Lasso-TD. Although verifying this condition is often infeasible (all the possible index sets S should be checked), in practice the empirical Gram matrix is likely to satisfy the compatibility condition (see van de Geer & Bühlmann 2009 for a detailed discussion and examples).

[2]The value function V^π is smooth in all the subsets of $\mathcal{X}$ where the policy π is constant, whereas its gradient is discontinuous in states where π changes.

Similar to Thm. 1, in order to prove Thm. 2, we first state and prove a Lemma about the Markov design model. This lemma provides a Lasso bound for the Markov design setting that has a fast rate, but requires some assumptions on the empirical Gram matrix.

Lemma 3. *We consider the model of regression with Markov design. Let $\Phi\hat{\alpha} \in \mathcal{F}_n$ be the Lasso prediction of the (noisy) values $Y = \{Y_t\}_1^n$, i.e., $\Phi\hat{\alpha} = \widehat{\Pi}_\lambda Y$. For any $\delta > 0$, with probability at least $1-\delta$, we have*

$$||f - f_{\hat{\alpha}}||_n \le ||f - f_\alpha||_n + \frac{6CL\sqrt{s}}{\psi}\sqrt{\frac{2\log(2d/\delta)}{n}}\,,$$

for any $\alpha \in \mathbb{R}^d$ such that the empirical Gram matrix $\frac{1}{n}\Phi^\top\Phi$ satisfies the $(3, S_\alpha)$-compatibility condition. Note that $s = s_\alpha$ and $\psi = \psi(3, S_\alpha)$.

Proof of Lemma 3. To simplify the notations, we write S instead of S_α. If the event $\mathcal{E}_1$ defined by Eq. 8 holds, for any $\lambda \ge 2\lambda_0$, Eq. 7 may be written as

$$||f - f_{\hat{\alpha}}||_n^2 + \lambda||\hat{\alpha}||_1 \le ||f - f_\alpha||_n^2 + \lambda||\alpha||_1 + \frac{\lambda}{2}||\hat{\alpha} - \alpha||_1\,. \quad (18)$$

By reordering of Eq. 18, we obtain

$$\begin{aligned} &||f - f_{\hat{\alpha}}||_n^2 + \frac{\lambda}{2}||\hat{\alpha}_{S^c}||_1 \le ||f - f_\alpha||_n^2 + \frac{\lambda}{2}||\hat{\alpha}_S - \alpha_S||_1 \\ &+ \lambda\big(||\alpha_S||_1 - ||\hat{\alpha}_S||_1\big) \le ||f - f_\alpha||_n^2 + \frac{3\lambda}{2}||\hat{\alpha}_S - \alpha_S||_1. \end{aligned} \quad (19)$$

Since for $||f - f_\alpha||_n \ge ||f - f_{\hat{\alpha}}||_n$ the statement of the lemma is true with probability one, we only consider the case where $||f - f_\alpha||_n \le ||f - f_{\hat{\alpha}}||_n$. In this case, we may deduce from Eq. 19 that $(\hat{\alpha} - \alpha) \in \mathcal{R}(3, S)$, and thus

$$||\hat{\alpha}_S - \alpha_S||_1^2 \le \frac{s||f_{\hat{\alpha}} - f_\alpha||_n^2}{\psi^2}\,. \quad (20)$$

Replacing $||\hat{\alpha}_S - \alpha_S||_1$ in Eq. 19 from Eq. 20 and using the triangle inequality, we have

$$||f - f_{\hat{\alpha}}||_n^2 \le ||f - f_\alpha||_n^2 + \frac{3\lambda\sqrt{s}}{2\psi}\big(||f - f_{\hat{\alpha}}||_n + ||f - f_\alpha||_n\big). \quad (21)$$

Solving Eq. 21 for $||f - f_{\hat{\alpha}}||_n$, we obtain

$$||f - f_{\hat{\alpha}}||_n \le ||f - f_\alpha||_n + \frac{3\sqrt{s}\lambda}{2\psi}\,. \quad (22)$$

The rest of the proof is similar to the end of the proof of Lemma 3, we first set $\lambda_0 = 2CL\sqrt{\frac{2\log(2d/\delta)}{n}}$ and $\lambda = 2\lambda_0$ that event $\mathcal{E}_1$ holds with probability $1-\delta$, and then plug in λ in Eq. 22. □

Proof of Theorem 2. Similar to Theorem 1, we first bound $||v - f_{\tilde{\alpha}}||_n$ as in Eq. 14, and then bound $||v - \widehat{\Pi}_\lambda \widehat{\mathcal{T}} v||_n$ using Lemma 3. □

5. Conclusions

In this paper, we analyzed the performance of Lasso-TD, a fixed point TD method in which the projection operator is defined as a Lasso problem. We first showed that this method is guaranteed to always have a unique fixed point and that its algorithmic implementation coincides with the LARS-TD (Kolter & Ng, 2009) and LC-TD (Johns et al., 2010) methods. We derived bounds on the prediction error of Lasso-TD in the Markov design setting, i.e., when the performance is evaluated on the same states used to train Lasso-TD. Although the first bound holds in a very general setting where no assumption is required, its estimation error has a slow decreasing rate w.r.t. the number of samples. At the cost of an extra assumption on the empirical Gram matrix, we also derived a sparsity oracle inequality which directly relates the prediction error to the sparsity of the value function in the feature space at hand. To the best of our knowledge, this is the first analysis of ℓ_1-regularized TD methods and is also the first result showing that Lasso-TD enjoys similar properties as Lasso in regression.

A number of open questions remains to be investigated: **1)** Derivation of bounds on the estimation of α^* when $v = f_{\alpha^*}$. This is the case where the value function lies in the function space, also known as the *high dimensional* assumption. **2)** Analysis of the sparsity of the Lasso-TD solution, i.e., $||\tilde{\alpha}||_0$. **3)** Extension to the random design setting, where the predictive performance is evaluated over the entire state space according to the stationary distribution of the policy.

Acknowledgements: We would like to thank Stéphane Gaiffas for providing us with a simpler proof for Lemma 1 and Odalric-Ambrym Maillard for useful discussions. This work was supported by French National Research Agency through the projects EXPLORA n° ANR-08-COSI-004 and LAMPADA n° ANR-09-EMER-007, and by PASCAL2 Pump Priming Programme on *Sparse RL in High Dimensions*.

References

Boyan, J. Least-squares temporal difference learning. *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 49–56, 1999.

Bradtke, S. and Barto, A. Linear least-squares algorithms for temporal difference learning. *Machine Learning*, 22:33–57, 1996.

Candes, E. and Tao, T. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

Farahmand, A. M., Ghavamzadeh, M., Szepesvári, Cs., and Mannor, S. Regularized policy iteration. In *Proceedings of Advances in Neural Information Processing Systems 21*, pp. 441–448, 2008.

Farahmand, A. M., Ghavamzadeh, M., Szepesvári, Cs., and Mannor, S. Regularized fitted Q-iteration for planning in continuous-space Markovian decision problems. In *Proceedings of the American Control Conference*, pp. 725–730, 2009.

Ghavamzadeh, M., Lazaric, A., Maillard, O., and Munos, R. LSTD with random projections. In *Proceedings of Advances in Neural Information Processing Systems 23*, pp. 721–729, 2010.

Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001.

Johns, J., Painter-Wakefield, C., and Parr, R. Linear complementarity for regularized policy evaluation and improvement. In *Proceedings of Advances in Neural Information Processing Systems 23*, pp. 1009–1017, 2010.

Kolter, Z. and Ng, A. Regularization and feature selection in least-squares temporal difference learning. In *Proceedings of the Twenty-Sixth International Conference on Machine Learning*, pp. 521–528, 2009.

Lazaric, A., Ghavamzadeh, M., and Munos, R. Finite-sample analysis of LSTD. In *Proceedings of the Twenty-Seventh International Conference on Machine Learning*, pp. 615–622, 2010.

Loth, M., Davy, M., and Preux, P. Sparse temporal difference learning using lasso. In *IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning*, pp. 352–359, 2007.

Mallat, S. *A Wavelet Tour of Signal Processing: Wavelet Analysis & Its Applications*. Academic Press, 1999.

Massart, P. and Meynet, C. An l1-oracle inequality for Lasso. Technical Report 00506446, INRIA, 2010.

Petrik, M., Taylor, G., Parr, R., and Zilberstein, S. Feature selection using regularization in approximate linear programs for Markov decision processes. In *Proceedings of the Twenty-Seventh International Conference on Machine Learning*, pp. 871–878, 2010.

Sutton, R. and Barto, A. *An Introduction to Reinforcement Learning*. MIT Press, 1998.

van de Geer, S. and Bühlmann, P. On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.